604160

ON STOCHASTIC LEARNING THEORY

Merrill M. Flood

P-353

19 December 1952

The RAND Corporation
1700 MAIN ST • SANTA MONICA • CALIFORNIA

ON STOCHASTIC LEARNING THEORY

Merrill M. Flood

1. Introduction

R. R. Bush and C. F. Mosteller[1]*, and also W. K. Estes[2], have proposed a stochastic theory of learning. They suppose that the organism makes a sequence of responses among a fixed finite set of alternatives, and that there is a probability $p_s(t)$ at moment t that response s will occur before moment t+1. They suppose further that the probabilities $p_s(t+1)$ are determined by: the $p_s(t)$, the response s_t actually made after moment t, and the event r_t that follows after response s_t. Specifically, they assume the functional form:

$$1.1) \qquad p(t+1) = M^{r_t s_t} p(t),$$

where p(t) is the m-dimensional vector whose s^{th} component is $p_s(t)$, and M^{rs} is a square stochastic matrix of order m whose elements M_{ij}^{rs} depend only upon r and s.

One especially interesting case is that in which there are just two classes of events; for example, they might be reward and non-reward. For the purposes of this paper, it will be sufficient to consider only this case since the more general case presents no added difficulty. Our object is to narrow the class of allowable matrices M^{rs} by making one more assumption that seems quite reasonable.

---

*Numbers refer to Bibliography at the end of the paper.

A number of suggestions have already been made for specializing the form of the matrices M^{rs}; several of these are discussed in a recent paper by R. R. Bush and G. L. Thompson,[5] both as to their mathematical form and as to their implications from the standpoint of learning theory. It is entirely possible that the proposal I am now making has already been considered but, if so, it has not come to my attention.

2. Background

I like to think of the elements of the matrices M^{rs} as physical constants that are characteristic of the organism, just as mass, color, and hardness may be thought of as physical constants pertaining to an object. The values of the M_{ij}^{rs} are to be estimated on the basis of data from an appropriate experiment, just as the mass of an object might be estimated from a set of observations taken during an experiment with a spring-balance. Furthermore, my theory is operationally well-defined only after I specify some single analytical process for estimating the constants, such as by averaging the observations.

This must be done by observing the organism in some situation where a correspondence is set up between a sequence of observed responses $\hat{s}_t$ and events $\hat{r}_t$ and the formal quantities s_t and r_t in the theory; the estimates $\hat{M}_{ij}^{rs}$ are necessarily functions of the observations $\hat{s}_t$ and $\hat{r}_t$, though perhaps a different one for each parameter. It does not matter that I am as yet unable to write these functions simply; all that is necessary is that there be a finite computational process that will yield the desired

estimates $\hat{M}_{ij}^{rs}$ in terms of the observations $\hat{s}_t$ and $\hat{r}_t$ for $t=1,2,\cdots,N$. To accomplish this, we first let $\hat{p}(t+1)$ be defined by the recursion relation

$$2.1) \qquad \hat{p}(t+1) = \hat{M}^{\hat{r}_t \hat{s}_t} \hat{p}(t) \quad \text{for } t=1,2,\cdots,N,$$

where $\hat{M}_{ij}^{rs}$ and $\hat{p}_k(1)$ are defined to be the values of the parameters M_{ij}^{rs} and $p_k(1)$ that maximize the "likelihood"

$$2.2) \qquad L[M_{ij}^{rs}, p_k(1)] = \prod_{t=1}^{N} \hat{p}_{\hat{s}_t}(t).$$

There is nothing new in what I have said so far, of course; these expository remarks simply provide the background for what follows.

3. Symmetry and Scope

If the theory is to be of much interest, and wide use, it must provide a valid description of a broad class of organismic behavior; the scope of the theory must be specified in terms of bounds for the class of behavioral situations explained by the theory. In particular, it should be possible to verify the theory by testing it for only some situations within a well-defined subclass before using it confidently for predictions concerning the remaining situations of the subclass. I believe this is what is meant by the term theory in scientific usage.*

As a special case, we might consider the subclass of all

---

* "In scientific usage, a HYPOTHESIS is a provisional conjecture regarding the causes or relations of certain phenomena; a THEORY is a hypothesis which has undergone verification, and which is applicable to a large number of related phenomena." Webster's New International Dictionary, Second Edition, Unabridged. G. and C. Merriam Company, Springfield, Mass., 1951, p. 2620.

two-choice situations for all animals. Still more specially, we might consider the subclass of all two-choice situations for some one human consisting at any one time of either: a) doing some particular thing (such as blinking), or b) not doing that thing. There is a serious problem in identifying this choice class at any one time with what appears to be the same situation at another time, but the important thing is to be able to do this so well that the theory does in fact check out closely and often; deviations between theoretical and observed behavior must then be sorted out after the fashion of the statistician and if after probing they seem to be unexplainable and satisfactorily small then the theory is considered to be valid for such purposes.

If we were lucky with the stochastic learning theory we might find a large class of human choice experiments explained by it in this sense. For instance, if the estimates of the parameters M_{ij}^{rs} were found to agree well in repeated trials with the same person in some one experimental learning situation then we would accept this as evidence that the $\hat{M}_{ij}^{rs}$ were physical constants and characteristic of the situation. If these same values were found for many types of humans, but another set of constants was found repeatedly for rats in the same experimental situation, then we would accept this happily as evidence that our theory had still wider scope. Thus, scientific development consists in increasing the extent of the subclass of situations that can be explained reliably by each hypothesis and in sharpening the boundary between this subclass and others in which the hypothesis fails.

And so with the stochastic learning theory. A good first step would be to find any experiment with humans, involving choices and rewards, that can be repeated over and over again yet always yielding essentially constant $\hat{M}^{rs}_{ij}$. Then a good second step would be confirmation of this constancy in quite a different experimental situation. As an example, suppose that the $\hat{M}^{rs}$ could be estimated reliably for a rat and a man with the same experimental design, just as they can both be weighed on the same scales, and that these estimates $\hat{M}^{rs}_{rat}$ and $\hat{M}^{rs}_{man}$ are used successfully to predict the amounts that the man and the rat would each win playing cooperatively in some carefully selected non-zero sum game; then this result would increase our confidence in the validity of the learning theory in such situations.

Turn now from scope to symmetry, and start with the notion that there must have been a first occurrence for each choice-situation met by the organism. On the first occurrence, essentially by definition, there would be no way for the organism to have a bias in favor of any single choice. Furthermore, if the theory is to be of use, the numbering of the choices is arbitrary and the validity of the theory cannot be dependent upon the numbering actually selected. In other words, the matrices M^{rs} must be such that any two vectors $q(t)$ and $q(t+1)$ obtained by applying the same permutation to the components of $p(t)$ and $p(t+1)$ must satisfy the relation

$$q(t+1) = M^{rs}q(t)$$

whenever

$$p(t+1) = M^{rs}p(t),$$

and provided that the permutation leaves the s^{th} component unchanged. In the next section, where this assumption is stated more precisely, it is shown that symmetry restricts the matrices M^{rs} very considerably; the number of independent parameters is reduced to three for each event class whenever there are more than two choices, and to two for each event class otherwise.

4. Symmetrical Model

We start with stochastic matrices M^{rs}, for r=0,1 and s=1,2,···,m, as in Section 1. In this section, it will sometimes be convenient to omit the superscript r when the argument is independent of this distinction. Furthermore, it will be enough to make the argument for some one value of s, say s=1, since an exactly similar argument holds for other values of s; so we also omit the superscript s, with the understanding that we are discussing only M^{r1} explicitly in this section, and consider the elements M_{ij} of this typical stochastic matrix $M=M^{r1}$.

Our symmetry assumption is now equivalent to the condition

4.1) $$TMp(t) = MTp(t),$$

where T is any permutation matrix such that $T_{11} = 1$ and where p(t) is any probability vector. Since p(t) can be any unit vector, an equivalent condition is simply that M must commute with every T. Thus we may set

4.2) $$TM - MT = \bar{M}.$$

An equivalent requirement is, therefore, that

$$4.3) \qquad \Gamma_{jx} = M_{ix} \quad \text{and} \quad \bar{M}_{yj} = M_{yi} \quad \text{for } i,j=2,\cdots,m; \quad x,y=1,2,\cdots,m.$$

It follows easily from 4.3), when $m > 2$, that M must be of the form

$$4.4) \qquad M^{r1} = \begin{Vmatrix} 1-(m-1)a^r & b^r & b^r & \cdots & b^r & b^r \\ a^r & c^r & & & & \\ a^r & & c^r & & \frac{1-b^r-c^r}{m-2} & \\ \vdots & & & \ddots & & \\ a^r & \frac{1-b^r-c^r}{m-2} & & & c^r & \\ a^r & & & & & c^r \end{Vmatrix}$$

where a, b, and c are three parameters subject only to the restrictions:

$$4.5) \qquad \begin{aligned} &0 \leq a \leq 1/(m-1) \\ &0 \leq b \leq 1, \\ &0 \leq c \leq 1, \\ &0 \leq b+c \leq 1. \end{aligned}$$

When m=2, there are only two parameters, and M is of the form

$$4.6) \qquad M = \begin{Vmatrix} 1-a & b \\ a & 1-b \end{Vmatrix},$$

where a and b are arbitrary within the closed unit interval. Of course, the general symmetrical model M^{rs} is obtained by permuting the 1^{st} and s^{th} rows and the 1^{st} and s^{th} columns of $M(a^r,b^r,c^r)$ as defined by the relations 4.4) and 4.6).

5. Some Comparisons

Bush and Mosteller[4] have proposed a specialization of the general operator M^{rs} of Section 1, called by them the "combining classes" model, which is determined by the requirement that condition 1.1) reduce to the condition

$$5.1) \qquad p_i(t+1) = a_i^{rs} p_i(t) + b_i^{rs}.$$

Of course, the parameters appearing in 5.1) are also subject to the restrictions

$$5.2) \qquad a_i^{rs} = 1 - \sum_{j=1}^{m} b_j^{rs} = a^{rs}, \quad \sum_{j=1}^{m} b_j^{rs} \leq 1 + \min\left(b_1^{rs}, b_2^{rs}, \cdots, b_m^{rs}\right),$$

$$0 \leq b_i^{rs} \leq 1.$$

In matrix form, as has been shown by Bush and Thompson,[3] 5.1) and 5.2) are equivalent to

$$5.3) \qquad M^{rs} = a^{rs} + (1-a^{rs}) \begin{Vmatrix} u_1^{rs} \\ u_2^{rs} \\ \vdots \\ u_m^{rs} \end{Vmatrix} \| 1 \; 1 \cdots 1 \; 1 \|;$$

where the u_i^{rs} are components of an arbitrary probability vector and a^{rs} is subject only to the restriction

$$1 \geq a^{rs} \geq \frac{\min_i u_i^{rs}}{\min_i u_i^{rs} - 1}.$$

It is obvious, upon comparison of 4.2) and 4.3), that the symmetrical model is of combining classes form if and only if $a^r = \frac{1-c^r-c^r}{m-2}$ in 4.4). On the other hand, the combining classes form 5.3) has symmetry if and only if a^{rs} and λ_s^{rs} are independent of s and

$$\lambda_1^{rs} = \frac{1-\lambda_s^{rs}}{m-1} \quad \text{if } 1 \neq s.$$

Thus, neither the symmetrical model nor the combining classes model can be obtained from the other by specializing parameter values. And, of course, the "symmetrical combining classes" model, defined by 4.4) and the condition that $a^r = \frac{1-c^r-c^r}{m-2}$, satisfies both the combining classes and symmetry assumptions.

Still more specialized models considered by Bush and Thompson[3], in which I shall be interested, include the two following, when written as special cases of 4.4);

Pure Model: $a^0=c^0=0$, $c^0=1$; $a^1=0$, $c^1+c^1=1$,

Mixed Model: $a^0=0$, $c^0=1-(m-2)a^0$, $a^1=0$, $c^1=(m-1)a^0$, $c^1=1-(m-1)a^0$.

5. The Error Problem

There is an important sense in which none of our models is yet suitable for experimental verification: we have made no allowance for observational error in the models. Although we have used the likelihood method (in Section 2) to provide an operational definition of our parameters, and even though the process with which we are dealing is stochastic, we still must

include some ordinary error parameters in our theoretical model before we can complete the estimation calculations in meaningful terms. I shall now introduce a new learning model, including an error parameter, to clarify this point.

As in Section 1, the component $p_1^*(t)$ of an m-dimensional probability vector $p^*(t)$ represents the probability of response 1 after moment t. Now we introduce an auxiliary m-dimensional probability vector $q(t)$, with components $q_1(t)$, that satisfies the stochastic relation

$$6.1) \qquad q(t+1) = M^{r_t s_t} q(t);$$

where M^{rs}, r_t, and s_t have the same definitions as in Section 1. Finally, we define $p^*(t)$ by the relations

$$6.2) \qquad p^*(t) = q(t) \quad \text{if } r_{t-1} = 0,$$
$$p^*(t) = e_{s_{t-1}} \quad \text{if } r_{t-1} = 1,$$

where e_x is the probability vector whose x^{th} component is unity; in words, this means that the organism never changes its choice after a rewarded response. It is immediately apparent, that the likelihood of 2.2) will be zero with this model for any set of observational data in which there is even a single exception to the rule "don't change on a winner;" and this will be the case even if the apparent exception is due to a clerical error.

We will modify the model by adding a new parameter θ representing the probability that at any moment t, after a rewarded choice s_{t-1}, that a choice $s_t \neq s_{t-1}$ will be made at random among the (m-1) alternatives rather than according to the

probability components of $p^*(t)$. This leads to a probability vector $p(t)$ defined by the relation

$$6.3) \qquad \begin{aligned} p(t) &= q(t) \quad \text{if } r_{t-1} = 0, \\ p(t) &= (1-\theta)e_{s_{t-1}} + \frac{\theta}{m-1}(J-e_{s_{t-1}}) \quad \text{if } r_{t-1} = 1, \end{aligned}$$

where J is the m-dimensional vector whose components are all unity. The new likelihood function is then

$$6.4) \qquad L[M_{ij}^{rs}, p_k(1), \theta] = \left[p_{s_1}(1) \prod_{t \neq t^*} p_{s_t}(t)\right]\left[(1-\theta)^A \left(\frac{\theta}{m-1}\right)^B\right],$$

where $r_{t^*-1} = 1$, A is the number of times $s_{t^*} = s_{t^*-1}$, and B is the number of times $s_{t^*} \neq s_{t^*-1}$. It is easily seen that the value of θ that maximizes this likelihood is

$$6.5) \qquad \theta = \frac{B}{A+B}.$$

With the statistical parameter θ included, a clerical error after a rewarded choice need no longer make the likelihood zero.

Perhaps a better way to include the error parameters is to distinguish between those occurring after a rewarded choice and those after a non-rewarded choice. This can be rationalized by noting that there are really two rather different types of errors, one class due to clerical and other observational flaws and another class due to the inexactness of the theory. For example, I suspect that changes after a rewarded choice that are made while the environment is really stationary, and with

no serial correlations. reflect the tendency for organisms to continually search for evidence of non-stationarity, and for serial correlation; none of our models is general enough to match such tendencies.

For a moment then, we will consider a model in which two statistical parameters appear. We start with any model, like those specified by 1.1) or 6.2), in which there is a vector $\bar{p}(t)$ whose component $\bar{p}_s(t)$ represents the approximate probability that the organism will make choice s after moment t. We next suppose that the probability is θ^1 after a rewarded choice, and θ^0 after a non-rewarded choice, that the organism will not choose according to the vector $\bar{p}(t)$; alternatively, the parameters θ^0 and θ^1 can each be thought of as joint probabilities of theoretical and observational error. The probability vector p(t) that we should use to represent total behavior is, therefore,

$$6.6) \qquad p(t) = \left[(1-\theta^1)\bar{p}(t)+\frac{\theta^1}{m}J\right]r_{t-1}+\left[(1-\theta^0)\bar{p}(t)+\frac{\theta^0}{m}J\right](1-r_{t-1}).$$

The likelihood estimates of θ^0 and θ^1, for this model, cannot be written explicitly for the general operator M^{rs} that yields $\bar{p}(t+1)$ from $\bar{p}(t)$; this was possible for the model of 6.3) only because the factors of the likelihood function involving θ were not dependent upon M_{ij}^{rs}.

7. A Preferred Model

I prefer the stochastic learning model defined by 6.3) and 6.6), where q(t) satisfies a stochastic relation of the form 1.1) with M^{rs} of the form 4.4) or 4.6). This model is not

of the form 1.1), where $M^{r_t s_t}$ is required to be a stochastic matrix, but it is of the form 1.1') if $M^{r_t s_t}$ is interpreted as a more general operator. This "Preferred Model" includes error parameters θ^j, provides directly for the "don't change on a winner" principle, and restricts the class of operator matrices as required by the symmetry assumption. The Preferred Model, for $m > 2$ and starting at moment t_o, has the following $m+6$ independent parameters: $p_i(t_o)$ for $i=1,2,\cdots,m-1$; a^j, b^j, c^j, θ^j for $j=1,2$. We think of the parameters θ^j as relevant to corrections both for clerical error, and for such theoretical errors as those due to apparent non-stationarity or serial correlation in the process.

The main hypothesis is that the theoretical scope for the Preferred Model is the class of all m-choice repetitive situations for organisms acting in a stationary serially uncorrelated environment, and one in which each repetitive m-choice situation has a first occurrence at moment t=1. The simplest experiments to test this hypothesis will probably be those in which the parameters $p_i(t_o)$ will all be taken equal to $(1/m)$, on the assumption that the experimental choice-situation represents a first occurrence, at least at the start of work with each new subject.

In a very strong sense, the main hypothesis would be supported if all the parameters were found to be essentially constant over a class of situations in which the value of m and the frequencies of reward and non-reward were varied widely. There would also be good support for the hypothesis if the

parameters a^J, b^J, and c^J were found to remain essentially constant over a class of situations in which m was held fixed, and the initial trial with each subject was controlled to be a first occurrence with $p_1(t_o) = (1/m)$, but the frequencies of reward and non-reward were varied.

Unfortunately, it is not at all necessary that any parameter explicit in the model be observationally rather constant, over an appropriate class of experimental situations; it would be enough if only certain functions of the parameters were observationally constant. In this sense, one can never reject the model; one can only note that a particular set of observations, as they were interpreted in terms of the model, do not lend support to the main hypothesis. Such a result is non-constructive, since progress requires success in observing constancies relative to the model. In all, then, the Preferred Model and the main hypothesis can never be more than guides for experimentation; in this sense, what we really have is only a guide to preferred experiments.

8. Some Preferred Experiments

There is much to be said for first testing the main hypothesis for the two-choice situation, since in this case there are two fewer parameters than when $m > 2$. And it will certainly be easier to control the starting vector $p(t_o)$ than to estimate it from the experimental data.

If success is met in the two-choice case, in the sense that all of the five parameters remain essentially constant,

then it will only be necessary to estimate the two new parameters in the three-choice case if it should be true that values are independent of m. If all this goes well, it will provide strong support for the theory if these same constant parameter values are found as m is increased.

It is reasonable to hope that all the parameters, in the m-choice case, will be essentially constant as the reward and non-reward frequencies are varied. In this favorable event, it will be desirable to try other variations in the conditions surrounding the m-choice case, such as amount of reward, in order to find the experimental bounds within which the model seems to be valid (its scope) in m-choice situations.

Perhaps the most important of all experimental design considerations is the requirement that the data not only be adequate to determine the parameters but that the estimation calculations be manageable. The best way that I can now see to handle the estimation problem is to keep the number of successive plays small in any one sequence after starting with a probability vector assumed known. More specifically if the total experiment with one subject consists of n sequences of N plays each, in which the starting vector in each sequence is (J/m), then N should be kept as small as possible and n should be fairly large. For example, if m=2 and N=4 then the parameters in all the models considered in this paper are easily estimated if n is large enough; the method of estimation is illustrated in Section 9.

This very brief outline can only suggest the direction for preferred experimentation at the start, since later designs must depend upon the results of earlier tests. Generally speaking the object is to proceed from the simple to the more complex as observed constancies permit this type of development.

9. Parameter Estimation

The general method of parameter estimation will be illustrated in this section by calculating some of the formulas for one model–experiment combination. The method is quite general, if the experiments are carefully designed for the purpose, and seems to provide manageable estimation formulas for most of the models discussed in this paper.

The model is defined by the following relations:

9.1) $$M^{r1} = \begin{pmatrix} a_r & b_r \\ 1-a_r & 1-b_r \end{pmatrix}, \qquad M^{r2} = \begin{pmatrix} 1-b_r & 1-a_r \\ b_r & a_r \end{pmatrix},$$

$$0 \le a_r \le 1, \quad 0 \le b_r \le 1.$$

9.2) $$q(t+1) = (\hat{r}_t M^{i\hat{s}_t} + (1-\hat{r}_t) M^{o\hat{s}_t}) q(t) \quad \text{for } t=1,2,\cdots,N.$$

9.3) $$p(t) = \hat{r}_t[(1-w)e_{\hat{s}_{t-1}} + w(J/2)] + (1-\hat{r}_t)q(t) \quad \text{for } t=2,3,\cdots,N.$$

9.4) $$p(1) = q(1) = (J/2).$$

The experimental design provides that:

a. Each of n trials requires five successive choices between two alternatives.

b. The probability of reward on the i^{th} choice in each trial is π_i, and the π_i are constant from trial to trial.

c. The result of reward or non-reward is in every instance independent of the choice actually made.

d. The pattern of reward and non-reward is determined independently for each trial by means of a table of random numbers.

In practice, I have used a punchboard with two columns and five rows. There are 32 possible patterns of reward and non-reward for one trial, since each of five rows is either rewarded or non-rewarded, and the choice among these is made according to a probability distribution determined by 32 probabilities φ_j; of course the π_i are determined uniquely by the φ_j. Most important of all, every effort is made to convince the subject that the trials are independent, so that he will not be influenced in his behavior in any one trial by his experiences in earlier trials. Actually, any other (5x2) design that preserves independence between trials, symmetry between columns, and independence between and within columns of one trial would provide the data necessary for estimation; provided all possible reward patterns appeared sufficiently often.

The basic observational quantities that we shall use are frequencies defined as follows:

9.5) $$F^{r_1 r_2 \ldots r_t}_{s_1 s_2 \ldots s_t}(s_{t+1}) = \text{Number of times in all } n \text{ trials that } s_{t+1} \text{ was chosen after prior choices } s_i \text{ and results } r_i.$$

We shall also need the associated quantities:

9.6) $F^{r_1 r_2 \cdots r_t}_{s_1 s_2 \cdots s_t}$ = Total number of times in all n trials that the choices were s_i and results r_i.

9.7) $$R^{r_1 r_2 \cdots r_t}_{s_1 s_2 \cdots s_t}(s) = \frac{F^{r_1 r_2 \cdots r_t}_{s_1 s_2 \cdots s_t}(s)}{F^{r_1 r_2 \cdots r_t}_{s_1 s_2 \cdots s_t}} .$$

The principle of the likelihood estimation method is to calculate theoretical probabilities $H^{\rho_t}_{\sigma_t}(s)$, as functions of some or all of the parameters of a model, corresponding to each of several independent observables $F^{\rho_t}_{\sigma_t}(s)$. Then the likelihood function $L_N = L_N(a_r, b_r, w)$, for n trials and N choices each trial, is defined by the relation:

9.8) $$\log L_N = \sum_{t=1}^{N} \sum_{\sigma_t, \rho_t} \sum_{s=1}^{2} F^{\rho_t}_{\sigma_t}(s) \log H^{\rho_t}_{\sigma_t}(s).$$

The required parameter estimates are the values of a_r, b_r, and w that maximize L_N, where each parameter is restricted to the closed unit-interval.

Unbiased estimates may also be obtained, though at a sacrifice of precision, by using any subproduct of L_N that includes the term $H^{\rho_t}_{\sigma_t}(2)$ if the term $H^{\rho_t}_{\sigma_t}(1)$ is present. This amounts to the assumption that the $H^{\rho_t}_{\sigma_t}(s)$ are independent distribution functions of s. It is sometimes convenient to make use of this selection principle, in order to simplify the numerical calculations, when the full set of parameters can thereby be broken into subsets for estimation purposes.

As an example, consider the case N=2; then

$$9.9)\quad \log L_2 = \sum_{r_1=0}^{1}\sum_{s_1,s=1}^{2} F_{s_1}^{r_1}(s)\log H_{s_1}^{r_1}(s) + \sum_{r_1,r_2=0}^{1}\sum_{s_1,s_2,s=1}^{2} F_{s_1s_2}^{r_1r_2}(s)\log H_{s_1s_2}^{r_1r_2}(s).$$

If only the terms involving $r_t=0$ and $s_2 \neq s_1=1$ are used, then just the two parameters a_o and b_o will appear, and the likelihood expression becomes

$$9.10)\quad \log L(a_o,b_o) = \sum_{s=1}^{2}[F_1^o(s)\log H_1^o(s)+F_{12}^{oo}(s)\log H_{12}^{oo}(s)],$$

where it is also assumed that $F_1^o(s) > 0$ and $F_{12}^{oo}(s) > 0$. We note that:

$$9.11)\quad 2H_1(1) = e_1'M^{o1}J = a_o+b_o,$$

$$9.12)\quad 2H_{12}^{oo}(1) = e_1'M^{o2}M^{o1}J = 1+(1-a_o-b_o)(1-a_o+b_o).$$

It follows easily, dropping the subscripts on a_o and b_o, that;

$$9.13)\quad \frac{\partial \log L}{\partial a} = \frac{F_1^o(1)}{a+b} - \frac{F_1^o(2)}{2-a-b} + \frac{(2a-2)F_{12}^{oo}(1)}{1+(1-a-b)(1-a+b)} - \frac{(2a-2)F_{12}^{oo}(2)}{1-(1-a-b)(1-a+b)}.$$

$$9.14)\quad \frac{\partial \log L}{\partial b} = \frac{F_1^o(1)}{a+b} - \frac{F_1^o(2)}{2-a-b} - \frac{2bF_{12}^{oo}(1)}{1+(1-a-b)(1-a+b)} + \frac{2bF_{12}^{oo}(2)}{1-(1-a-b)(1-a+b)}.$$

If we set

$$\frac{\partial \log L}{\partial a} = \frac{\partial \log L}{\partial b} = 0,$$

and solve for a and b with the conditions $R_1^o(1) \neq R_1^o(2)$, $R_{12}^{oo}(1) \neq R_{12}^{oo}(2)$, and $a+b \neq 1$, we obtain

9.15) $$\frac{R_i^{o}(1)}{a+b} = \frac{1-R_i^{o}(1)}{2-a-b}\ , \quad \text{and}$$

9.16) $$\frac{R_{i2}^{oo}(1)}{1+(1-a-b)(1-a+b)} = \frac{1-R_{i2}^{oo}(1)}{1-(1-a-b)(1-a+b)}\ ;$$

whence

9.17) $$a = \frac{2[R_i^{o}(1)]^2 + R_{i2}^{oo}(1) - 1}{2R_i^{o}(1) - 1}\ ,$$

9.18) $$b = \frac{2[R_i^{o}(1)]^2 - 2R_i^{o}(1) - R_{i2}^{oo}(1) + 1}{2R_i^{o}(1) - 1}\ .$$

If either $R_i^{o}(1) = R_i^{o}(2)$ or $R_{i2}^{oo}(1) = R_{i2}^{oo}(2)$ then all solutions of 9.15) and 9.16) satisfy the condition $a + b = 1$, and we shall consider this case separately. We are left with the result that the function L_2, defined by 9.10), does not attain its maximum value within the restricted range of the parameters a_o and b_o except possibly when:

a. a_o and b_o satisfy 9.17) and 9.18) and lie within the closed unit-interval, or

b. $a_o = 0$, or

c. $b_o = 0$, or

d. $a_o = 1$, or

e. $b_o = 1$, or

f. $a_o + b_o = 1$.

It is easily seen, remembering that $P_i^{o}(s) > 0$, that

$$L_2(0,0) = L_2(1,1) = 0,$$

$$L_2(a_o,1-a_o) = 2^{-(F_1^o+F_{12}^{oo})} > 0,$$

$$L_2(a_o,0) = 2^{-(F_1^o+F_{12}^{oo})} a_o^{F_1^o(1)+F_{12}^{oo}(2)} (2-a_o)^{F_1^o(2)+F_{12}^{oo}(2)} (1+[1-a_o]^2)^{F}$$

and that $L_2(a_o,0) > L_2(a_o,1-a_o)$ for some value of a_o in the open unit-interval. Consequently, it follows that L_2 attains its maximum value, within the restricted range for a_o and b_o, only if:

9.19)

a. a_o and b_o satisfy 9.17) and 9.18), or

b. $a_o = 0$ and b_o is in the open unit-interval, or

c. $b_o = 0$, and a_o is in the open unit-interval, or

d. $a_o = 1$ and b_o is in the open unit-interval, or

e. $b_o = 1$ and a_o is in the open unit-interval.

It is a tedious but straightforward calculation to determine the estimated values with the help of the conditions 9.19).

In one pilot experiment, the observed values were

$$F_1^o(1) = 2, \quad F_1^o(2) = 80, \quad F_{12}^{oo}(1) = 41, \quad F_{12}^{oo}(2) = 19,$$

and so

$$R_1^o(1) = 1/41 \text{ and } R_{12}^{oo}(1) = 41/80.$$

Using 9.19), we have

$$b_o = \frac{\frac{2}{41} - \frac{2}{41} - \frac{41}{80} + 1}{\frac{2}{41} - 1} < 0.$$

Consequently, we must consider the possibilities:

$$
\begin{aligned}
2^{142}L(0,b) &= b^{40}(2-b)^{80}(2-b^2)^{41} = H_1(b),\\
2^{142}L(a,0) &= a^{21}(2-a)^{99}(2-2a+a^2)^{41} = H_2(a),\\
2^{142}L(1,b) &= (1+b)^{43}(1-b)^{121}(1+b^2)^{19} = H_3(b),\\
2^{142}L(A,1) &= (1+a)^{2}(1-a)^{162}(1+2a-a^2)^{19} = H_4(a).
\end{aligned}
\tag{9.20}
$$

The following inequalities are easily verified:

$$H_4(a) < 2^{21}, \quad H_3(b) < 2^{62}, \quad H_2(1/5) > 10^{19} > 2^{62},$$

and the required solution is therefore determined by $H_1(b)$ or $H_2(a)$. The maximum values of $H_1(b)$ and $H_2(a)$ are approximately

$$H_2(1/5) = 10^{19} > H_1(2/3) = 10^{11}$$

and so the required estimate is

$$\hat{a}_o = 0.220, \quad \hat{b}_o = 0;$$

it is also to be expected that neither $\hat{a}_o$ nor $\hat{b}_o$ is 1 because 9.15) requires that their sum be approximately 2/41. The estimate for $\hat{a}_o$ can be determined more exactly by solving for it as a root of the equation

$$101a_o^3 - 264a_o^2 + 244a_o - 42 = 0. \tag{9.21}$$

We turn now to the observables $F_1^1(s)$ and $F_{11}^{11}(s)$. The likelihood function corresponding to these quantities is

$$L(w) = (1-w/2)^{F_1^1(1)+F_{11}^{11}(1)}(w/2)^{F_1^1(2)+F_{11}^{11}(2)}.$$

It is easily seen that the value of w that maximizes L(w) is

$$w = 2\,\frac{F_i^{I}(2) + F_{ii}^{II}(2)}{F_i^{I} + F_{ii}^{II}}.$$

More generally, if all the experimental data were used to estimate w, it is easily seen that the result is:

$$w = \frac{2f^*}{f+f^*},$$

where f is the number of repeats after a winner and f^* is the number of non-repeats after winners.

For the pilot experiment the estimate is

$$w = \frac{2f^*}{f+f^*} = \frac{0}{33}, \quad \text{using } N = 3 \text{ only.}$$

Actually, in the pilot experiment, the subject always repeated after a winner so the estimate is still w = 0 when all the data are used.

The estimation of a_1 and b_1 requires that data for N = 4 be used, and the calculations are a bit more tedious so the details are omitted here. The method is exactly analogous to that just used for obtaining estimates of a_o and b_o.

10. Summary

A stochastic learning model is proposed in which:

a. Explicit provision is made for errors of observations.

b. Separate allowance is made for the "don't change on a winner" principle.

c. The number of independent paramters is reduced, from that in the general matrix operators used by Bush and Mosteller, by a symmetry assumption.

d. A preferred model is introduced and discussed, but it is not likely that the learning theory represented by this model will have great scope.

e. The significance of these matters, with respect to measurement of the physical constants hypothesized by such stochastic learning models, is discussed in relation to similar questions pertaining to a few alternative models.

f. An experiment is described, together with a method for estimating parameters, that should be adequate to provide a critical test of the various alternative theories discussed.

BIBLIOGRAPHY

1. Bush, R. R., and Mosteller, C. F., "A Mathematical Model for Simple Learning." Psychological Review, Vol. 58, No. 5, Sept. 1951, pp. 313-323.

2. Estes, W. K., "Toward a Statistical Theory of Learning." Psychological Review, Vol. 57, No. 2, March 1950, pp. 94-107.

3. Bush, R. R., and Thompson, G. L., "Multiple Choice Situations." Working Memoranda 26 and 28, University of Michigan Summer Seminar on Design of Experiments on Decision Processes, Santa Monica, June-July 1952.